ARI Research Note 90-81

Developing Training and Evaluation Scenarios for Armor Using Simulation Networking—Developmental (SIMNET-D)

Robert S. DuBois and Joseph A. Birt

Universal Energy Systems, Inc.

for

Contracting Officer's Representative
Barbara A. Black

Field Unit at Fort Knox, Kentucky
Donald F. Haggard, Chief

Training Research Laboratory
Jack H. Hiller, Director

July 1990

United States Army
Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

JON W. BLADES
COL, IN
Commanding

Research accomplished under contract
for the Department of the Army

Universal Energy Systems, Inc.

Technical review by

Carl W. Lickteig

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS --
2a. SECURITY CLASSIFICATION AUTHORITY -- 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE --	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S) UES 788-031	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Research Note 90-81
6a. NAME OF PERFORMING ORGANIZATION Universal Energy Systems, Inc. 6b. OFFICE SYMBOL (If applicable) --	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute Fort Knox Field Unit
6c. ADDRESS (City, State, and ZIP Code) Human Factors Division 4401 Dayton-Xenia Road Dayton, OH 45432-1894	7b. ADDRESS (City, State, and ZIP Code) ATTN: PERI-IK Fort Knox, KY 40121-5620
8a. NAME OF FUNDING / SPONSORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences 8b. OFFICE SYMBOL (If applicable) PERI-I	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F4168986 D0052
8c. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 63007A PROJECT NO. 795 TASK NO. (351)3101 WORK UNIT ACCESSION NO. C2

11. TITLE (Include Security Classification)
Developing Training and Evaluation Scenarios for Armor Using Simulation Networking-Developmental (SIMNET-D)

12. PERSONAL AUTHOR(S)
Du Bois, Robert S.; and Birt, Joseph A. (Universal Energy Systems, Inc.)

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 88/06 TO 89/12	1990, July	41

16. SUPPLEMENTARY NOTATION
Barbara A. Black, Contracting Officer's Representative

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Simulation Networking-Developmental (SIMNET-D) M1 tank Block II

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report documents a series of procedures and guidelines for developing training and evaluation scenarios or exercises using the Army's interactive simulation testbed, Simulation Networking-Developmental (SIMNET-D). These procedures result from and are illustrated by the experiences gained through the conduct and support of concept evaluations of the M1 Abrams Block II subsystem using SIMNET-D.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Barbara A. Black	22b. TELEPHONE (Include Area Code) (502) 624-2613 22c. OFFICE SYMBOL PERI-IK

ACKNOWLEDGMENTS

The authors wish to acknowledge the support of research scientists and technical staff of Universal Energy Systems (UES) and the U.S. Army Research Institute for the Behavioral and Social Sciences at Fort Knox (ARI-KNOX). The contributions of these individuals spanned the SIMNET-D training and evaluation experiences that form the basis for this research. Mr. Paul Smith, a former UES research scientist, assisted in the development and administration of scenarios in support of the POSNAV, IVIS, CITV, and C^3 efforts. LTC Daniel Motola (USA, Ret.) provided necessary Armor subject matter expert (SME) support throughout all scenario generation efforts. Dr. Robert Howells provided editorial support. Ms. Terry Kerr, Ms. Kristi Kyne, Ms. Donna Cullen, and Ms. Colleen Crisp provided administrative support. Several research assistants, including Mr. Glenn Exum, Ms. Alicia Sawyer, Ms. Cordette Herrington-Swain, and Ms. Rhonda Thomason, provided valuable and professional assistance with soldier training, scenario administration, and data collection activities. Drs. Barbara Black, Carl Lickteig, and Kathy Quinkert of ARI-KNOX provided valuable technical guidance throughout all UES SIMNET-D contract experiences.

DEVELOPING TRAINING AND EVALUATION SCENARIOS FOR ARMOR USING SIMULATION NETWORKING-DEVELOPMENTAL (SIMNET-D)

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CONTENTS (Continued)

DEVELOPING TRAINING AND EVALUATION SCENARIOS FOR ARMOR USING SIMULATION NETWORKING-DEVELOPMENTAL (SIMNET-D)

The Army's Simulation Networking-Developmental (SIMNET-D) is a very recent and advanced interactive simulation test bed for combined arms systems. Until a series of M1 Abrams Block II subsystem concept evaluations was conducted using SIMNET-D, there was insufficient experience with the facility to develop meaningful guidelines for developing SIMNET-D based training and evaluation scenarios. Based on these M1 evaluation experiences, however, this note provides a set of procedures for developing Armor training and evaluation scenarios for use in SIMNET-D. These guidelines are intended to support future users of the SIMNET-D facility such as combat development officers recently assigned to perform SIMNET-D evaluations, researchers new to SIMNET-D, and subject matter experts (SMEs) assigned to scenario development. This document cannot be considered a substitute for the operations and research expertise of these team members. Rather, it should help focus this expertise on the cumulative and cooperative front-end efforts required for a successful evaluation.

Background

SIMNET-D entails some unique capabilities and constraints as a test bed for Armor training and evaluation activities. Below is a description of the SIMNET-D test bed, as well as a review of the previous M1 Abrams Block II evaluations that provided the experience base for developing these scenario generation guidelines.

The SIMNET-D Experimental Test Bed

General Description

The Army's SIMNET-D test bed interactively links a variety of combined arms simulators, including M1 tanks, Bradley Fighting Vehicles, Forward Area Air Defense (FAAD) vehicles, and generic fixed and rotary-wing aircraft. In addition, SIMNET-D includes microcomputers representing tactical, administrative, and logistical combat service support elements (Bolt, Beranek, & Newman (BBN) Laboratories, 1986; Miller & Chung, 1987; Perceptronics, Inc., 1986). SIMNET-D can support both local-area (within site) and long-haul (across site) network interaction of simulators for combined arms battalion and below research. SIMNET-D's combat simulators and relevant research capabilities, including their advantages and disadvantages, have been described in previous U.S. Army Research Institute (ARI) technical reports (Du Bois & Smith, in preparation; Du Bois & Smith, 1989). A brief description of the SIMNET-D test bed resources is included.

SIMNET Combat Vehicle Simulators

SIMNET vehicle simulators model real system behavior to the minimum degree necessary for soldiers to perceive them as realistic and acceptable (Chung, Dickens, O'Toole, & Chiang, 1988; U.S. Army Armor School (USAARMS), 1987). Consistent with a selective fidelity design, however, SIMNET vehicles do not include all weapon system components. For example, the SIMNET M1 lacks the real M1's machine guns, auxiliary sight, and open-hatch. Likewise, the simulators' visual systems only present daylight environments. Figures 1 and 2 depict the current M1 module's driver's compartment and turret area, respectively.

Individual simulators are supported by a terrain and operations database and by audio and visual systems for modeling battlefield conditions, equipment status, and weapon system performance. All simulator battlefield locations, positions, status, control activities, and weapon effects are linked and updated across an Ethernet.

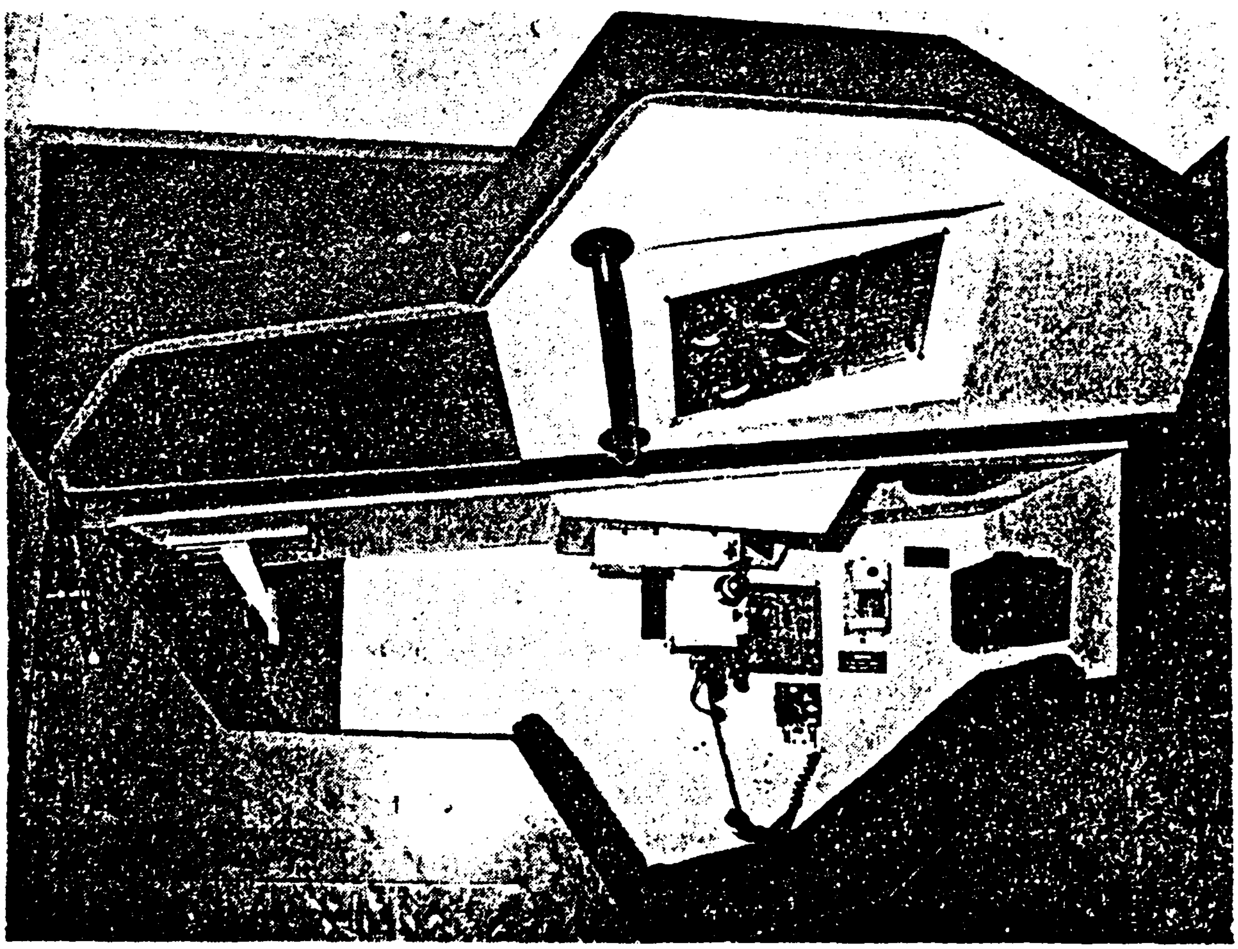

Figure 1. The M1 simulator's driver's compartment.

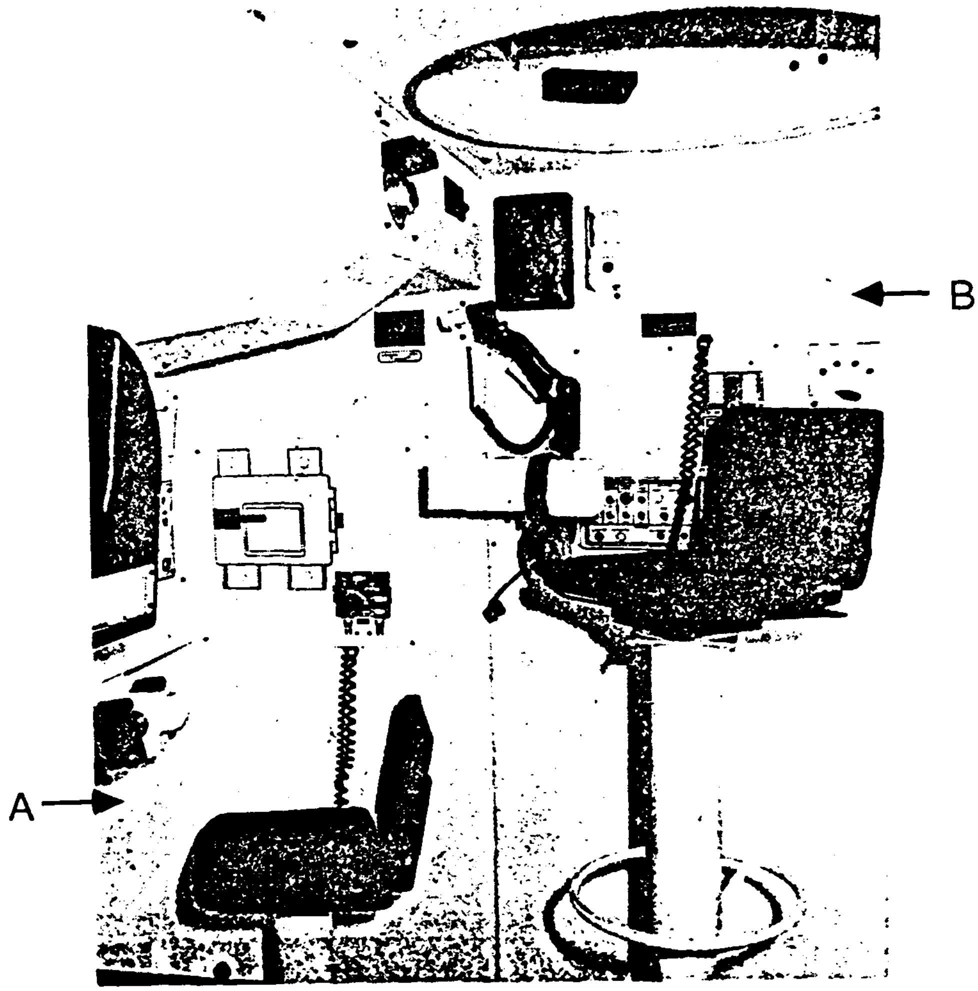

Figure 2. The M1 simulator's turret area, including (A) the gunner's stations and (B) the tank commander's station. The loader's station is not shown.

Consistent with the test bed philosophy, developers continuously modify the capabilities of SIMNET-D simulators. Recently, for example, a thermal imaging capability was added to the SIMNET-D M1 modules to support an evaluation of the Commander's Independent Thermal Viewer (CITV). The CITV is a target acquisition subsystem proposed for the Block II M1 Abrams tank. In each evaluation case, researchers must thoroughly examine the current features of both the simulated module(s) and the combat development subsystem(s) to ensure that their level of fidelity is appropriate with respect to the training and research issues to be evaluated.

SIMNET-D Training and Research Features

The SIMNET-D test bed directly supports several training and research features, including: (a) reconfigurable simulators; (b) Semi-Automated Forces (SAFOR) workstations; (c) Plan View Displays (PVDs); (d) shadow view monitors; (e) the Management, Command and Control System (MCC); and (f) the Data Collection and Analysis System (DCA). These features are described in Table 1.

SIMNET-D Disadvantages and Advantages

Disadvantages. The SIMNET-D environment currently has several primary limitations when compared with field operational settings, these are: (a) the closed hatch only nature of vehicle simulators, (b) the minimal visual cues presented (i.e., low fidelity computer imagery graphics), and (c) the minimal attention to motion cues. Also, like their field counterparts, SIMNET-D equipment (particularly the simulators, SAFOR, and any simulated vehicle subsystems to be evaluated) does break down.

The potential for these limitations to affect training and evaluation findings must be carefully weighed. For example, research has demonstrated significant degradations in tank crew and platoon closed-hatch navigation and target acquisition in the field (Barron, Lutz, Degelo, Havens, Talley, Smith, & Walters, 1976). Similar degradation can be expected in the closed-hatch simulator when compared to open-hatch field performance. Closed-hatch limitations, however, may serve as a positive caveat for some training and research activities. Most AirLand Battlefield operations will occur in closed-hatch, nuclear, biological, chemical (NBC), artillery and small arms fire environments. Hence, SIMNET-D results may generalize quite well to the future battlefield.

SIMNET-D does offer compensatory features for reducing the effects of some of these limitations. For example, the M1 tank simulator includes an Azimuth Indicator, Turret Reference System, and special paper maps. SIMNET-D combat scenarios can also be carefully designed to control limitation effects on exercise performance. For example, to control the effect that limited SIMNET-D visual cues may have on navigation, discrete terrain locations can be selected to serve as the operational setting. In addition, participant training programs can focus on specific procedures for working around these limitations (e.g., review map-based navigation techniques such as polar plot, intersection and resection). Nevertheless, until research is conducted to assess the validity of SIMNET-D for training and research, evaluators must be careful in assuming that their effects generalize to field and, ultimately, to actual combat performance.

Table 1

Description of SIMNET-D Training and Research Features

Features	Description
Reconfigurable Simulators	The SIMNET-D simulator hardware and software are reconfigurable. Hence, the Army can simulate, evaluate, and redesign a new capability, like a new combat vehicle or subsystem, before building the actual system.
Semi-Automated Forces (SAFOR)	The SAFOR is a multi-vehicle simulation program for creating and controlling automated, unmanned, opposing and friendly forces' aircraft and vehicles.
Plan View Display (PVD)	The PVD monitor provides a "bird's eye view", in real time or playback, of a SIMNET-D scenario. The PVD depicts a terrain map and provides map manipulation and event flagging functions.
Shadow View Monitors	Shadow view monitors allow trainers and experimenters to observe, in real time or playback, SIMNET-D scenario events from selected vehicle vision blocks and sights.
Management, Command and Control System (MCC)	The MCC provides service support stations and functions for battle management, simulator and target placement, fire support, close air support, and combat service support.
Data Collection and Analysis System (DCA)	The DCA supports automated soldier performance measurement. The DCA includes the Data Logger (DL), RS/Probe[1] (previously DataProbe), and RS/1. The DL records all Ethernet data packet traffic. RS/Probe and RS/1 are data management and analysis software packages.

[1]"RS/Probe", "DataProbe', and "RS/1" are registered trademarks of BBN Software Products Corporation.

Advantages. SIMNET-D offers some unique advantages over other simulations or field exercises. For example, the fidelity of collective task performance in SIMNET-D may be greater with respect to the realism of task-loaded environments, the realism of combat stress levels, and the capability for automated, objective performance measurement (Du Bois & Smith, in preparation; Du Bois & Smith, 1989). In SIMNET-D, soldiers can execute collective tasks not supported by other simulations. SIMNET-D also supports performance measurement at battalion levels and below. SIMNET-D allows evaluators to collect diverse data, including mission, soldier, training, organization, doctrine, and human factors (e.g., soldier-machine-interface) measures. Many of these measures can not be objectively, safely, or economically evaluated in the field.

Previous ARI Supported SIMNET-D Research

To support the development and evaluation of the modernized M1 Abrams tank, the ARI Field Unit at Fort Knox has conducted evaluations of selected Block II M1 components. These components include: (a) an automated navigation system, the Position Navigation system (POSNAV); (b) an automated command, control, and communication (C^3) system, the Intervehicular Information System (IVIS); and (c) an advanced target acquisition system and sighting capability, the Commander's Independent Thermal Viewer (CITV). In addition, as part of the IVIS evaluation, Du Bois (1989) documented the need for and initial development of simulation-based C^3 performance assessment methods for Armor small unit commanders.

These evaluations (POSNAV, IVIS, CITV, and C^3) form the experience base for outlining Armor scenario development guidelines for SIMNET-D. Each Block II experiment required the development of collective Armor platoon scenarios for the training and evaluation of tank crew and platoon performance. For the POSNAV evaluation (Du Bois & Smith, 1989), the scenarios emphasized navigation performance. For the IVIS evaluation (Du Bois & Smith, in preparation; Du Bois, in preparation), the scenarios emphasized C^3 performance. For the CITV evaluation (Quinkert, in preparation), the scenarios emphasized gunnery and C^3 performance. For all evaluations, the Armor scenarios were developed and administered by researchers from Universal Energy Systems, Inc., an experimental support contractor for ARI.

Scenario Development Procedures

The scenario generation procedures used in ARI SIMNET-D evaluations to date included seven distinct steps. These steps are shown in Figure 3. These steps include: (a) define the objectives of the research and/or training, (b) select the evaluation and scenario strategy, (c) draft the general scenario requirements, (d) iteratively redesign the scenario (using shakedown evaluations), (e) produce the detailed "final" draft scenario, (f) conduct the pilot evaluation, and (g) finalize the

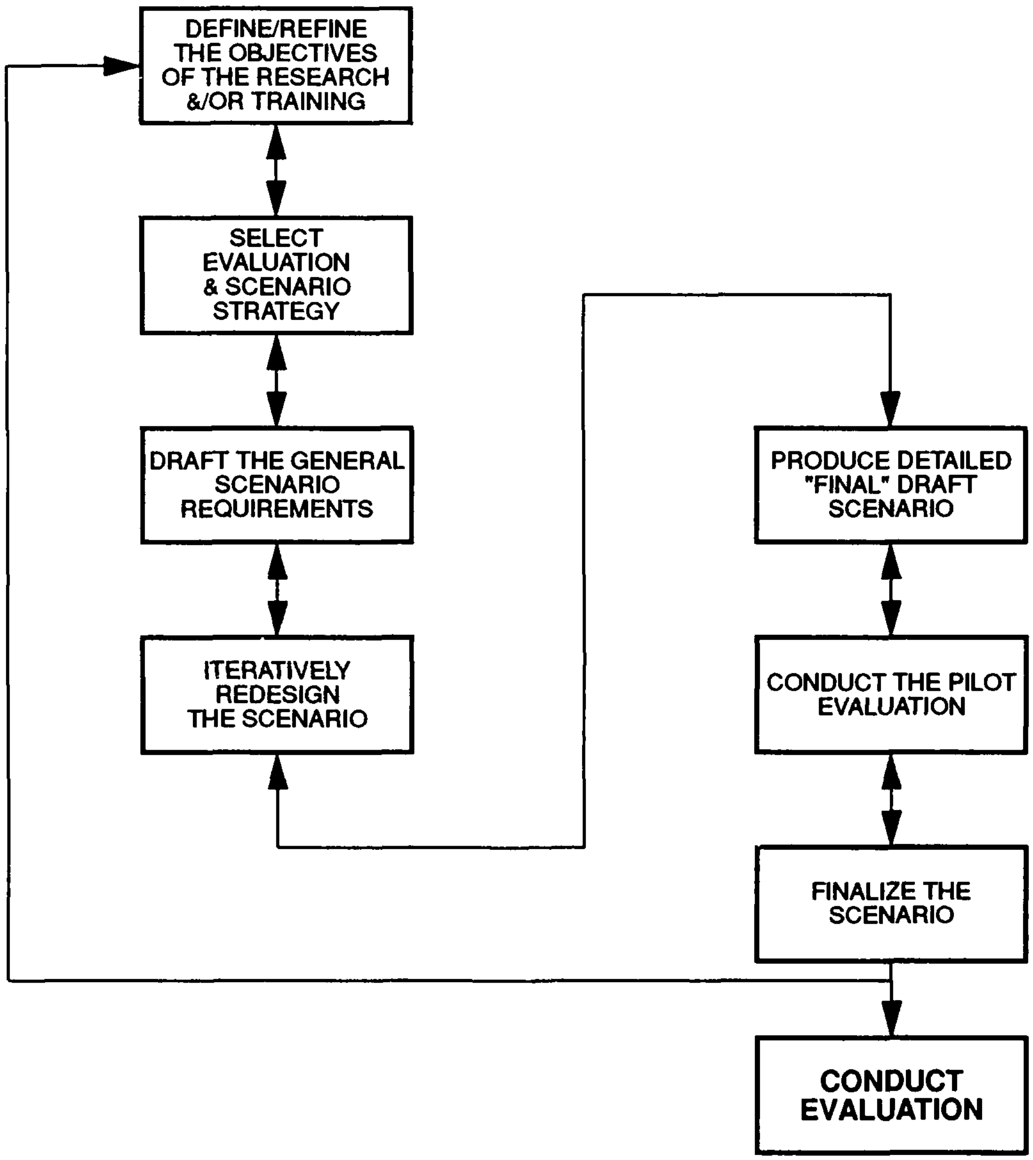

Figure 3. Seven major steps for developing SIMNET-D scenarios.

scenario. In addition, to illustrate the importance of feedback throughout scenario development, feedback mechanisms are shown among steps. For example, information acquired during scenario development can have implications for redefining the ultimate research or training objectives.

Define the Objectives of the Research/Training

The first step in the development of training and evaluation scenarios using SIMNET-D is to define and prioritize the objectives or goals of the research or training. In most cases, these objectives will already be known, but will probably only be defined in very general or global terms. Evaluators may be assigned to evaluate a new piece of equipment without detailed direction or clearly defined objectives (e.g., evaluate the Block III tank). Likewise, trainers may be assigned to develop and evaluate a training strategy used to introduce a new piece of equipment. The initial objective of the POSNAV and IVIS evaluations, cited earlier, was to identify the soldier performance requirements of these new systems. For the purposes of designing scenarios, such objectives have to be refined.

The refined objectives that result from this step should be systematically derived through an iterative analysis. As Figure 4 shows, one begins with the sponsor's statement of the problem. As just noted, the sponsor's objective may be quite nonspecific.

Initial questions are then asked of the sponsor to further identify a prioritized list of evaluation objective issues for the investigation at hand. The evaluator uses a set of questions that are specific to the sponsor's problem and that address the results, expectations, and purposes of the investigation under consideration. The research problem may focus upon one or more objective evaluation areas such as:

* the combat or operational performance effects of new hardware/software systems or subsystems

* the relative effectiveness of alternative designs or user interfaces among system components including the human component

* the impact of doctrinal or organizational changes to operations

* the effectiveness of a training change or innovation

* the impact of a manpower or personnel change

The sponsor may have focused upon the most important of these areas; still, questions about all of the areas should be asked. Even though the sponsor may be interested in exploring or covering all areas and contingencies, the evaluator should determine which high priority areas will have the most impact on future decisions. With limited evaluation and simulation resources, the evaluators must devise the most meaningful and

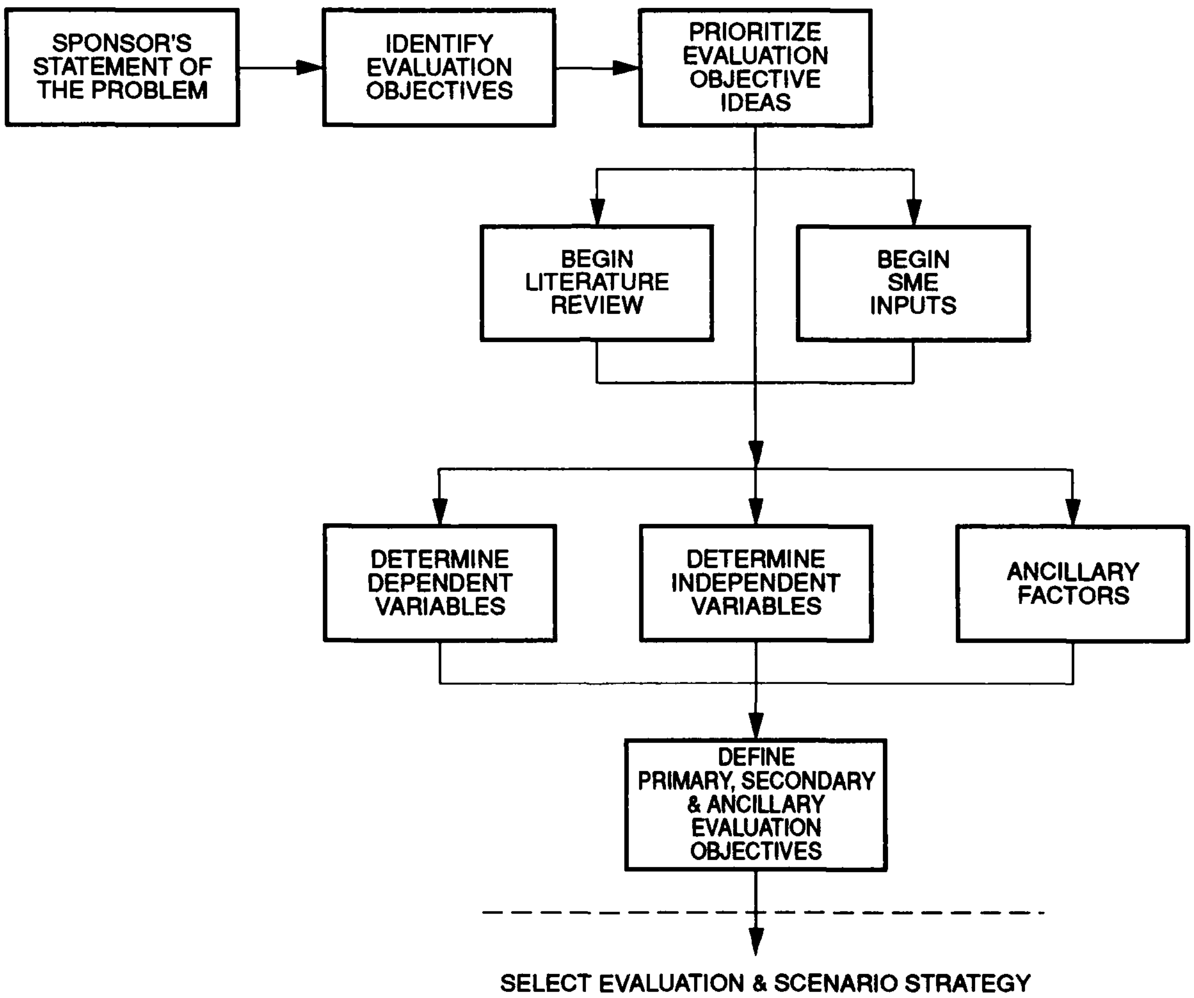

Figure 4. Scenario development: steps for defining the evaluation objectives.

feasible experimental strategies, hypotheses, and plans. Even with the substantial data collection and analysis (DCA) resources, it is not advisable to take on multipurpose research that simultaneously addresses potentially confounding research problem areas. Objectives for many research areas often call for unique experimental paradigms.

In the case of the POSNAV and IVIS investigations, early interactions with the sponsors focused upon the soldier performance implications of new subsystem concepts. Although it was of interest to learn as much as possible about the POSNAV interface and training implications, as well as any potential doctrinal impacts, these were not the primary focus of the initial experimental planning. Further interaction revealed that one interface feature of the POSNAV map display (grid matrix presentation only versus grid and terrain feature presentation) did warrant experimental treatment. Other areas, such as training, were primarily controlled so they did not have a major impact upon the primary measures. General information was collected via questionnaire items about other interface features and about training experiences and needs.

With the prioritized research objective areas in hand, an ongoing literature review can clarify the evaluation issues and procedures. Previous research reports and Army manuals can provide important information on critical training and evaluation concerns and methods for addressing them. In addition, reviews of relevant Army doctrine descriptions, task analyses, and training exercises will help evaluators identify potential scenario missions, tasks, and measures, as well as the doctrinal principles that guide them.

Concurrent with the literature review process, it is beneficial to arrange and begin systematic interaction with subject matter experts (SMEs). These interactions are of utmost importance to scenario development. SME selection should be based on the research and training issues of concern. For example, for the Block II subsystem evaluations, discussions with Armor combat developers, active unit personnel, school instructors, and test officers resulted in a clearer set of objectives with respect to current Armor concerns, system objectives, evaluation needs, doctrine, potential scenario missions, tasks and measures, and SIMNET-D resources and requirements.

If the SMEs come from the sponsoring organization, these SME discussions not only produce important information and insights, but also initiate a positive environment for future acceptance of evaluation procedures and findings. SME requirements occur throughout the entire scenario generation process. Hence, evaluators are urged to identify an SME team during this stage. This team should actively participate across all scenario development and evaluation phases.

As a result of the prioritized objectives, the SME interactions, and the literature review, evaluators should determine the generic dependent and independent variables and ancillary factors that need to be considered for their investigation. Dependent variables represent the performance, training, human factors, or other categories of measures that will be evaluated. Independent variables represent the particular experimental or training factors of interest.

At this early stage, dependent and independent variables are defined as general categories of measures and controls. The specific measures and factors to be evaluated will most likely be scenario dependent or the result of final subsystem configuration or specifications. For example, specific gunnery measures addressed will depend on the number of targets engaged, their placement, and any specific composite scoring strategies used. While general gunnery measure categories can be defined (e.g., target acquisition, engagement accuracy, hit rate), specific data capture requirements and scoring schemes must relate to specific scenario events.

The selection of measurement categories and strategies to meet stated objectives does impact upon and interact with scenario development. For instance, when selecting measures at the scenario level, one may consider composite measures within or across scenarios or one may even select scenarios of varying levels of difficulty for measurement and interpretation purposes. As one proceeds with this process, one should keep in mind the important relationships between scenario selection and interpretability of the measure and of the experiment itself.

At this point, the evaluator is looking for categories of dependent variables that can be convincingly linked to primary problem areas. General measure selection should not be based on simply collecting those measures already provided by existing DCA capabilities (e.g., fuel usage, distance travelled). Evaluators should determine how individual and collective performance measures, mission and system effectiveness measures, safety, training, workload, and other human factors criteria may satisfy the major goals of the evaluation.

Independent variables will be associated with the primary research purposes which the evaluator is trying to state clearly in objective terms. In the case of the POSNAV and IVIS evaluations, soldier performance was observed with and without the subsystems of interest and, in the POSNAV case, with an alternative interface configuration of the map display.

In general, for each category of evaluation objective, there will be different sets of independent variables. If research objectives include training innovations, manipulations of independent variables may include some pretraining conditions and some posttraining conditions. For example, when using SIMNET-D to evaluate a proposed embedded training package, trainers may compare the posttraining performance of soldiers who were trained using conventional procedures with the performance of similar soldiers who were trained using the embedded training package.

For many experimental purposes, it is also possible to think of scenarios themselves as independent variable conditions. In most cases, when scenarios are involved it is because the evaluator is trying to generalize or predict outcomes for present or future operational possibilities. Thus, scenario production may be seen as random sampling from all possible operational conditions or as a fixed selection among cases of interest, such as degraded mode, continuous operations, or intense conflict. When more than one scenario is relevant, they may be regarded as levels or categories of an independent variable. For example, in the IVIS evaluation both an offensive and defensive future battlefield scenario were selected to evaluate the subsystem's platoon combat performance effects.

At this point, the evaluator can better state the primary, secondary, and ancillary research objectives in priority order and in terms of the generic variables and factors of importance in the investigation. For instance, the evaluation objectives of the POSNAV investigation were as follows:

* POSNAV: Primary Objective: To compare the navigation performance of Armor crews and platoons using POSNAV with the navigation performance of Armor crews and platcons using conventional tools and procedures.

Secondary Objective: To compare the navigation performance of Armor crews and platoon using a grid matrix POSNAV map display with the navigation performance of Armor crews and platoons using a terrain feature POSNAV map display.

Ancillary Objectives: To evaluate soldier reactions to the POSNAV soldier-machine-interface and recommend design changes. To evaluate soldier reactions to the prototype POSNAV training program and recommend future POSNAV training guidelines. Based upon mission performance and soldier reactions, identify potential impacts of POSNAV on Armor personnel selection, manpower considerations, and doctrine.

Select the Evaluation and Scenario Strategy

Supported by more specific evaluation objectives, literature review findings, and SME insights, as discussed above, the evaluator is now prepared to identify an evaluation and scenario strategy. The process of selecting these strategies is outlined in Figure 5.

First, the evaluator should identify alternative evaluation and scenario strategies which could be used to meet the evaluation objectives. These strategies should be linked to the general categories of independent and dependent variables identified earlier. Alternative evaluation and scenario

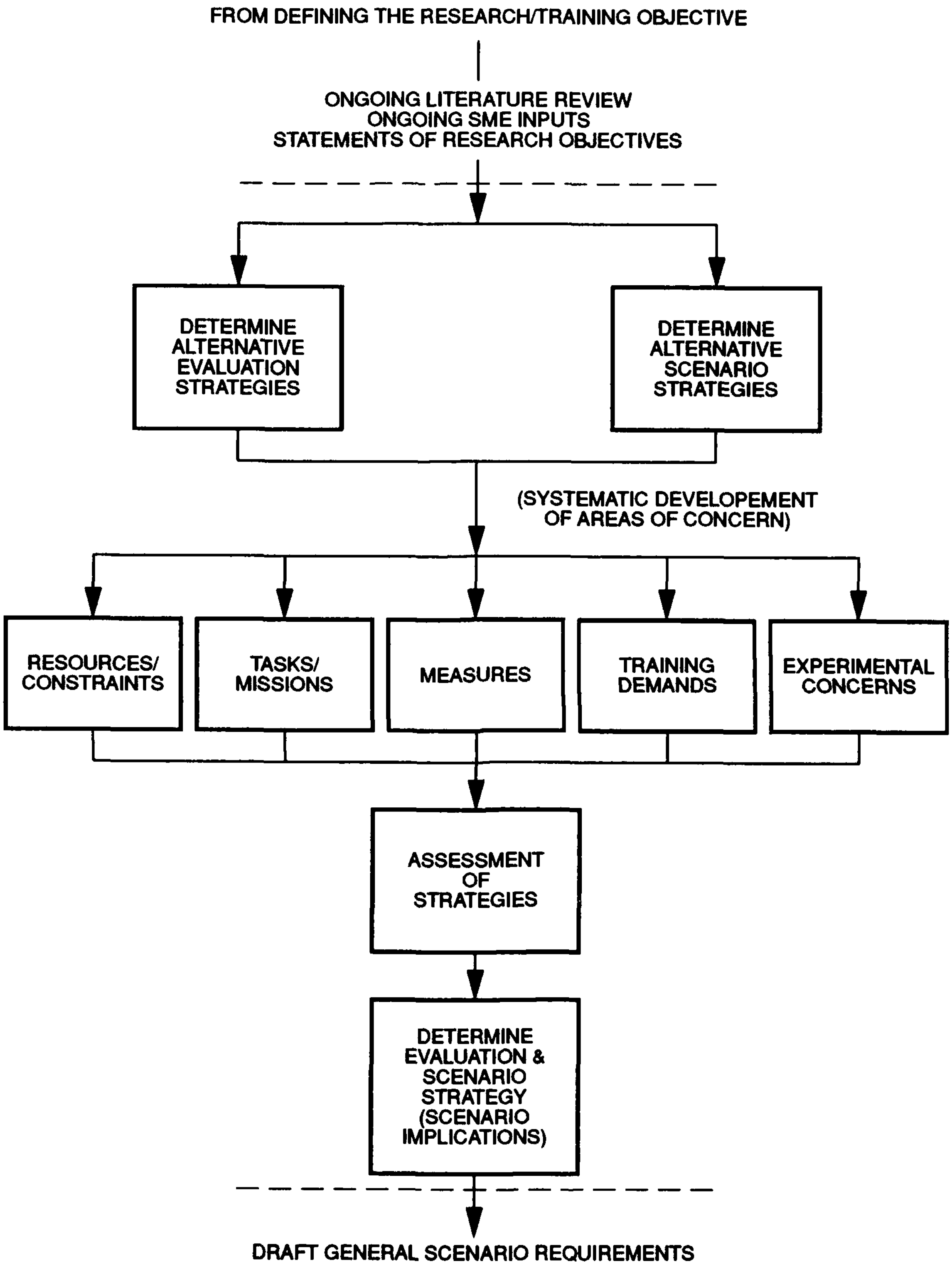

Figure 5. Scenario development: steps for selecting the evaluation and scenario strategy.

strategies can be those (a) used in previous research, (b) identified by SMEs, or (c) resulting from "brainstorming" interviews or creative interactions with the researcher or program originator. In any case, the evaluator will need to tailor or create them for the specific investigation at hand.

Evaluation strategies refer to the experimental approach used for conducting the research or training evaluation and consideration for the order of the evaluation activities. Evaluation strategies specify how one will treat the independent and dependent variables, as well as the ancillary factors. For example, in the POSNAV, IVIS, and CITV evaluations, alternative evaluation approaches included both within and among-group designs.

Within-group designs allow for each unique group of evaluation participants to complete evaluation scenarios under all treatment conditions (e.g., with and without the new piece of equipment). Among-group designs require each group of participants to represent only one treatment condition (e.g., conduct the scenarios either with or without the new piece of equipment). Each of these types of evaluation strategies, as well as mixed designs which include both within and among factors, are associated with unique advantages and disadvantages. For example, within-group designs are more statistically powerful (allow for more confidence in one's findings) and may require fewer groups of soldiers. However, within-group designs can require more evaluation time per group. It is not our purpose to treat experimental design for research in depth in this paper. Excellent books and courses exist to address this subject (e.g., Keppel, 1982; Winer, 1968).

While scenario developers should not have to design experiments, they should be prepared to work closely with evaluation designers. The evaluation strategy or strategies chosen for addressing the research or training objectives result in parameters for scenario development (e.g., length of scenarios, number of scenarios, and type of scenarios applicable).

Scenario strategies refer to the specific approach used to generate data for addressing the evaluation objectives. One uses scenarios in an evaluation as a systematic means of capturing the relevant training or operational contexts for the performance in question. In simulation, one may consider the scenario to be representative of the real world contexts and activities of interest. Since it may be difficult to impossible to sample from all likely contexts, the evaluator may adopt an approach to develop a typical, "a realistic", the most expected, or a worst case scenario, depending upon the evaluation objectives. The scenario conditions and activities may be considered to include the psychological sets provided for the performance being measured. If multiple scenarios are used to cover the contexts of interest (e.g., offensive, defensive), then scenarios may be considered an independent variable for experimental comparisons.

To assess soldier performance, for example, evaluators could use several strategies ranging from requiring full mission performance to requiring only select task performances. This range of possible scenarios is only limited by the imagination of the evaluator and the practical constraints in reaching experimental answers to the research problems. For example, a common scenario requires performance of a combat mission. Some factors of a mission scenario which evaluators could vary include:

* Degree of Participant Freedom (e.g., free play versus rigid experimenter control).

* Friendly Unit Capabilities (e.g., complete capabilities versus only selected vehicles or resources represented).

* Enemy Unit Capabilities (e.g., manned simulator-based force versus SAFOR vehicles or MCC targets; shooting versus non-shooting force).

* Mission Completeness (e.g., partial versus complete mission performance required).

* Mission Conditions (e.g., degraded versus normal status).

* Doctrine (e.g., future battlefield context versus current battlefield context).

When using full mission scenarios, evaluators should divide the mission into several unique mission segments. For example, specific mission events, such as a phase line crossing, fragmentary order transmission, or battle position occupation, could be used to identify the beginning of a new mission segment. By identifying mission segments, evaluators provide a basis for estimating the percent of each mission completed by soldiers. Moreover, in the case of a major equipment breakdown, evaluators can restart the exercise at the last segment start point. In addition, mission segments can provide unique mini-exercises used to collect repeated measures of selected soldier performances.

Complete free play mission performance is not necessarily required for SIMNET-D scenarios. In some cases, evaluators may find it more appropriate to simply develop scenarios which require the repeated performance of only select mission segments or combat tasks. For example, to evaluate crew target acquisition and gunnery performance with or without CITV, gunnery target exercises similar to those used with the Conduct of Fire Trainer (COFT) were developed (Quinkert, in preparation). To evaluate small unit C^3 performance with or without IVIS, a single tank tactical exercise was developed that required repeated performance of nine C^3 tasks (Du Bois, 1989).

The decision concerning the selection of the appropriateness of any evaluation and scenario strategies should be based on evaluation objectives, task requirements, simulation capabilities, unit demands, and evaluation constraints. The kinds of strategies available in the free-play SIMNET-D environment are limited primarily by evaluator imagination. For some training or evaluation issues such as embedded training or soldier-machine-interface design, actual simulator modules may not be necessary during actual data collection. The use of SIMNET-D voice recording capabilities, PVDs and SAFOR work stations permits evaluators to examine commander C^3 behaviors in responses to previously recorded or staged combat exercises.

Once evaluators have identified a pool of potential evaluation and scenario strategies, they are ready to begin the process of carefully examining the requirements, advantages, and disadvantages of each. Five unique factors or features of each approach which should be identified and evaluated are: (a) the resource requirements and constraints, (b) the tasks and/or missions supported, (c) the measures supported, (d) the training demands, and (e) the experimental concerns. Below is a description of each of these factors.

Resource Requirements and Constraints. For each evaluation and scenario strategy, evaluators should carefully consider the known or probable evaluation constraints. Of particular importance are time, troop, and resource constraints. Evaluators should consider the number of test support personnel and soldiers available to participate in the evaluation, as well as the availability of necessary SIMNET-D resources (e.g., PVDs, SAFOR, and simulators).

These estimates can have critical implications for scenario development, including the type, duration, and number of scenarios that can be conducted. Final troop and other support requirements will have to be identified and approved for Army support well in advance of the proposed evaluation dates.

In addition to troop, time, and resource constraints, evaluators should identify any other demands or requirements for conducting the scenario in the SIMNET-D environment. Based on the specific mission and/or tasks identified, evaluators should determine the friendly and threat capabilities (e.g., vehicles, size, competency) necessary for scenario execution.

Tasks and/or Missions Supported. Based upon the literature review, discussions with SMEs, simulation system limitations, and other constraints, evaluators should identify the potential tasks related to their primary objective. For example, in the POSNAV, IVIS, and CITV evaluations, tasks were chosen from the land navigation, C^3, and gunnery domains, respectively. Specific crew and platoon tasks identified for these evaluations were those deemed compatible with performance and assessment in SIMNET-D and those potentially affected by the introduction of future tank subsystems.

Evaluators should consider each evaluation and scenario approach with respect to the degree to which they support the execution of relevant tasks and missions. For example, free play and force-on-force full mission scenarios may allow for execution of many global unit tasks, but could provide few discrete soldier task performance assessment opportunities. Moreover, free play mission approaches may not allow for repeated, objective task measures or valid inferences as to their cause.

For training purposes, evaluators may have to match scenario and evaluation strategies to those tasks for which soldiers are deemed potentially deficient. For example, evaluators may be charged with training and evaluating the C^3 performance of Armor company commanders based on poor Army Training and Evaluation Program (ARTEP) performance or based on unit request for training (e.g., sustainment training). In such cases, evaluators should identify approaches suited for the assessment of critical C^3 tasks.

Measures Supported. Evaluation and scenario approaches can differ on their ability to capture relevant categories of data/measures. Evaluators should consider the capability of each approach to allow for reliable, accurate, and objective assessment. The resources required to collect such measures from each approach should also be weighed.

Training Demands. SIMNET-D presents soldiers with some unique training needs for both research and training applications. Before actual evaluation and scenario approaches can be selected, evaluators should consider the training and preparation requirements necessary to support each approach. These training demands could include the need to ensure that soldiers become familiar with the differences between the actual or prototype system and the simulated capabilities. Moreover, training for specific Armor scenario requirements (e.g., land navigation, C^3, and gunnery) may be necessary to ensure that soldiers understand the scenario execution requirements and the measures to be obtained. This is especially important if scenario approaches require significant deviations from standard operating procedures and require a high degree of experimenter control.

Training requirements should be developed with respect to the experience and education levels expected of the troops to be assigned to support the evaluation. For example, if soldiers are expected from all experience levels or from non-active Armor units, evaluators should be sure to include relevant refresher training. When troop support availability requires the use of incomplete units (e.g., three men crews), evaluators should be sure to consider additional administrator and soldier training requirements.

Experimental Concerns. As considered in the preceding discussion of among and within-groups designs, evaluation and scenario approaches can vary on the degree to which inferences can be confidently made from findings. For example, evaluation approaches can differ in the degree to which they are internally and externally valid. While these issues are of principal concern to the experimenter (and not the scenario developer), the importance of these features must be emphasized. The scenario developer should be appraised of the degree to which the evaluation and scenario approaches allow for experimental confidence and integrity.

Based upon the consideration of each of the five features discussed above, evaluators should select the approach or approaches best suited for the research problems of interest. Approach decisions should result from an individual and group decision making process involving SMEs, scenario developers, test support personnel, and the evaluation originators. The basic implications of the final evaluation/scenario strategies for the scheduling of activities, troop support, administration, training, resource needs, and the tasks and/or missions and measures supported should be documented.

Draft the General Scenario Requirements

Once evaluators have selected an evaluation and scenario strategy, they should begin to draft the scenario(s). The requirements for this step are illustrated in Figure 6 and are described below.

Identify Terrain Requirements. Based on the type of mission and/or tasks chosen to be included in exercise scenarios, evaluators should choose an area of terrain on the SIMNET-D battlefield suitable for scenario execution. For example, the POSNAV evaluation focused specifically on land navigation performance. Hence, for platoon evaluations, offensive missions were selected and were placed on expansive cross-country terrain locations void of significant, discrete terrain features. For the IVIS and CITV platoon evaluations, however, the scenarios were designed to ease navigation requirements and demand intensive C^3 and gunnery performance, respectively. Hence, more discrete offensive and defensive locations were chosen, which included easily recognizable terrain features and control point locations.

Evaluators should be careful in selecting terrain areas to ensure that they provide for adequate cover and concealment and movement for mission or task accomplishment. Several areas on the SIMNET-D battlefield are cluttered with uncrossable rivers and can present problems for scenario event placement. Moreover, only a few areas on the SIMNET-D battlefield include simulated hull and turret down positions. Terrain areas also differ with respect to the degree they offer in-tank soldiers with line-of-sight. SIMNET-D paper maps and PVD terrain maps should be carefully reviewed to identify appropriate terrain areas for SIMNET-D scenarios.

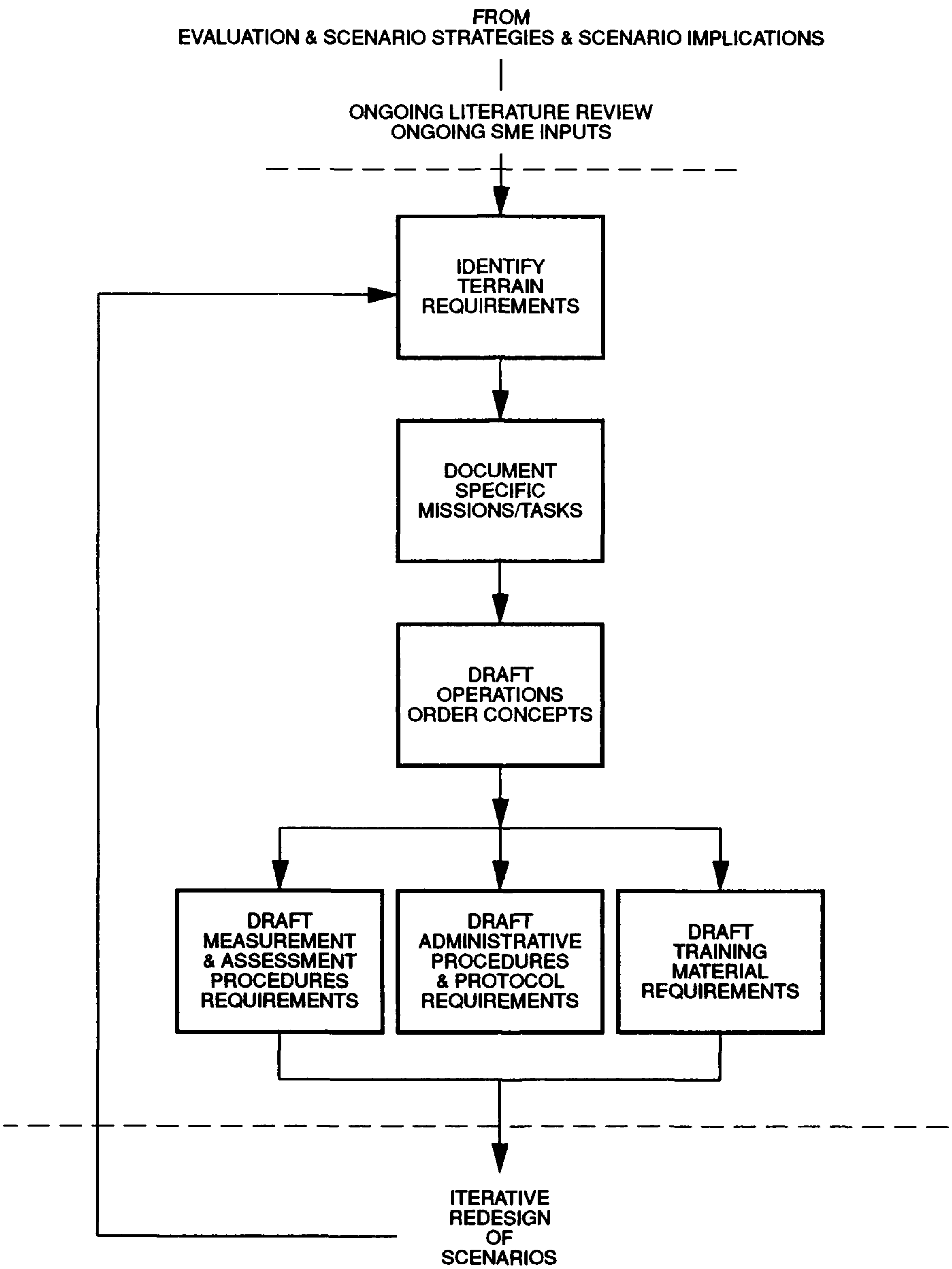

Figure 6. Scenario development: steps for drafting the general scenario requirements.

Document Specific Missions and/or Tasks. After having identified appropriate terrain locations for scenario placement and execution, the evaluators should direct their efforts toward the specification of missions and/or tasks along the battlefield. Evaluators should be primarily concerned with the identification of draft locations for task accomplishment, not in documenting the final scenario. Hence, the best advice is to use a non-permanent marker when marking on any paper map overlays. In addition, based upon the specific tasks chosen, evaluators should document repeated opportunities for task performance and assessment.

In determining locations for specific scenario enemy activity and target engagements, evaluators should consider the line of sight or intervisibility offered to scenario participants. Target placements should be selected that provide effectively performing units an opportunity for acquisition and engagement. PVD and SAFOR line-of-sight estimates can form the initial basis for target placements, but should be confirmed by actually getting in the simulator and examining the terrain.

Adequate terrain and time should also separate specific scenario tasks, to allow for performance assessment for each event. Repeated back-to-back scenario events, while often expected in actual combat, may not allow for standardization and performance assessment. An element of experimental control of battlefield activity is recommended to ensure appropriate assessment, especially when automated measurements are complemented by controller or administrator logs.

While placing tasks, evaluators should also carefully document all locations and orientations for targets, as well as all event locations and requirements. Any SAFOR vehicle locations, routes, and parameters should be carefully determined and stored as discrete hard disk files. All specified file labels should be easily recognizable and provide for quick recognition and execution.

Draft Operations Order Concepts. When reviewing the draft mission and task locations and requirements, as well as doctrine, evaluators should state the specific concepts that will guide the scenario operation. The concepts should describe the friendly and enemy situations, mission, artillery fire placements, and any combat service or close air support that will be provided. The operations order should be as detailed as necessary to ensure that soldiers understand the scenario requirements.

Throughout this step, evaluators should not try to "reinvent the wheel," avoiding the duplication of any materials used in previous and relevant efforts. If the Armor school or units have proven operations orders and local area (e.g., Fort Knox) terrain locations that could be adapted for SIMNET-D use, they should be reviewed. For example, for the POSNAV, IVIS, and CITV evaluations, general operation order concepts, terrain locations, and overlays from Armor School and Command and Staff scenarios served as the skeleton for draft scenario development. By adapting already approved and proven scenarios to SIMNET-D use, evaluators also improve the face validity or eventual Army commander and soldier acceptance of research and training procedures used and evaluation findings.

Document Desired Measures and Assessment Procedures. Based upon the general categories of measures (dependent variables) identified earlier, evaluators should identify and document the measures that they want to collect and the procedures they intend to use to collect them. The selection of specific measures for evaluation in SIMNET-D should, as with scenario tasks, be based on the ultimate objectives of the research. Evaluators should try to keep from falling into the "collect everything" frame of reference that is so easy to establish in SIMNET-D.

In the field, evaluators often are limited to only battle outcome measures or off-tank observer ratings. In SIMNET-D, however, evaluators can choose from a nearly limitless array of objective measures, as well as in-tank, off-tank, shadow view monitor-based and PVD-based ratings. Furthermore, voice recording capabilities also support the collection of radio traffic from multiple radio networks for later review and analysis. Stealth view capabilities, available at AIRNET now and which will eventually be part of SIMNET-D, also allow researchers to view exercise events in real time or playback from any vehicle or battlefield location, as well as any elevation. These capabilities should be fully understood by evaluators before measures are selected.

Previous research efforts, both field and simulation-based evaluations, as well as Army task analyses, doctrine, and SME judgments can be especially helpful in determining measures. Gound and Schwab (1988) for example, provide tables which rate the degree to which selected Armor Army Training and Evaluation Program (ARTEP) tasks can be performed in SIMNET-D. Unfortunately, however, most task analyses often include subjective performance standards, if any. In these cases, evaluators and SMEs should estimate these standards. Du Bois (1989) outlines procedures for developing both criterion-oriented (standard-based) and normative-oriented (absolute measurement-based) composite measures for small unit C^3 task performance.

All automated measurement requirements should be briefed to SIMNET-D data collection and analysis personnel to determine whether current resources can be used to obtain these measures. Moreover, significant time is often required to develop DataProbe and RS/1 collection and analysis routines. The earlier that SIMNET-D analysts can begin to evaluate specific research and training measurement requirements, the better. If data collection routines require test administrator or in-tank observer procedures, evaluators can also develop training plans for ensuring accurate assessment.

Draft Administration Procedures and Protocols. Specific scenario administration procedures and protocols should be drafted. Any administrator scripts or materials necessary for controlling scenario execution should be drafted. These scripts should include a coding system to allow each exercise administrator and data collector to consistently identify each scenario segment.

MCC target and simulator initialization files and PVD overlays and flagging menus should also be drafted. PVD scenes which allow for the most effective overwatch of exercise activities should be identified and saved as discrete overlay files. These overlays should be developed using the PVD in as much detail as necessary to allow for standardized administration and assessment. MCC files for exercise initialization and any support requirements should be outlined. As with SAFOR file labels, MCC and PVD file labels should be simple, yet easily recognizable and distinguishable.

Administrator personnel requirements and responsibilities should also be identified. For example, administrators should be identified as to who will operate the PVD, SAFOR, radio, and MCC stations during each scenario.

Draft Training Materials. It is important to examine soldier and test support training needs across all phases of scenario development. In fact, test support personnel (e.g., administrator, trainer, and data collector) training can often be accelerated by including these personnel in the scenario development process.

At this stage of scenario development, training materials should be drafted to support the scenario redesign and pilot evaluation process. Checklists should indicate the tasks selected for training. The training tasks should be linked with training approaches, including lecture, videotape, hands-on individual and collective practice, and training scenario execution. Training evaluation materials, used to assess the effectiveness of soldier training, should also be developed. Potential training evaluation materials can include knowledge tests, performance evaluations, and questionnaires. At a minimum, evaluators should prepare group discussion material to evaluate the effectiveness of, and soldier reactions to, the training program.

For actual training efforts, including the training program evaluation, a draft training plan should accompany the draft scenarios. This plan should describe the draft training process, including an estimated time line, necessary materials and resources, and administrative and measurement procedures.

Iteratively Redesign the Scenario

Once scenarios have been initially drafted on SIMNET-D terrain and documented as described earlier, considerable time and effort should be expended to iteratively redesign and refine the exercises. This initial draft evaluation and refinement process is shown in Figure 7 and is described below.

Shakedown Evaluations. Shakedown evaluations should be conducted in simulation using test support personnel, SMEs, and active Armor soldiers, if possible. Moreover, the PVD can be used to evaluate the intervisibility between locations across the scenario terrain and the SAFOR can provide a dry run exercise practice capability for reviewing mission enemy and friendly situations and requirements. Shakedowns are especially important for examining specific scenario task placements and battlefield locations, data collection procedures, training procedures, and administration requirements.

Evaluate Task Placement and Requirements. As part of the iterative scenario refinement process, evaluators should consider the placement of tasks and their execution requirements. Events often appear quite logical and salient on paper, but do not work well in simulation. Evaluators can identify the potential responses of units to scenario events and determine necessary changes to ensure that appropriate task responses are possible and realistic.

Task requirements should be repeated throughout the scenario to cover as many task conditions as possible. For example, target placement for gunnery task performance in the CITV evaluation were designed to vary target engagement ranges, unit sizes and locations, vehicle status (e.g., moving or stationary), and types (trucks, personnel carriers, tanks). In addition, friendly vehicles were placed along the battlefield to evaluate the occurrence of fratricide. By repeating task performance, evaluators improve their ability to reliably or confidently assess task performance.

In the evaluation of task placements, practice or shakedown runs in simulation are especially important to ensure that task initiating stimuli are salient and produce desired responses in a realistic manner. For example, target activities should be visible to oncoming units and should appear realistic.

Respond to Unique Simulation Constraints/Features. SIMNET-D terrain presents some unique task constraints which should be addressed during scenario refinement. The SIMNET-D battlefield, for example, includes few realistic terrain areas for cover and concealment, as well as hull and turret down locations. Hence, task placements and mission requirements should be identified which reduce the effects of these constraints. Administration procedures and operations concepts can also be adapted.

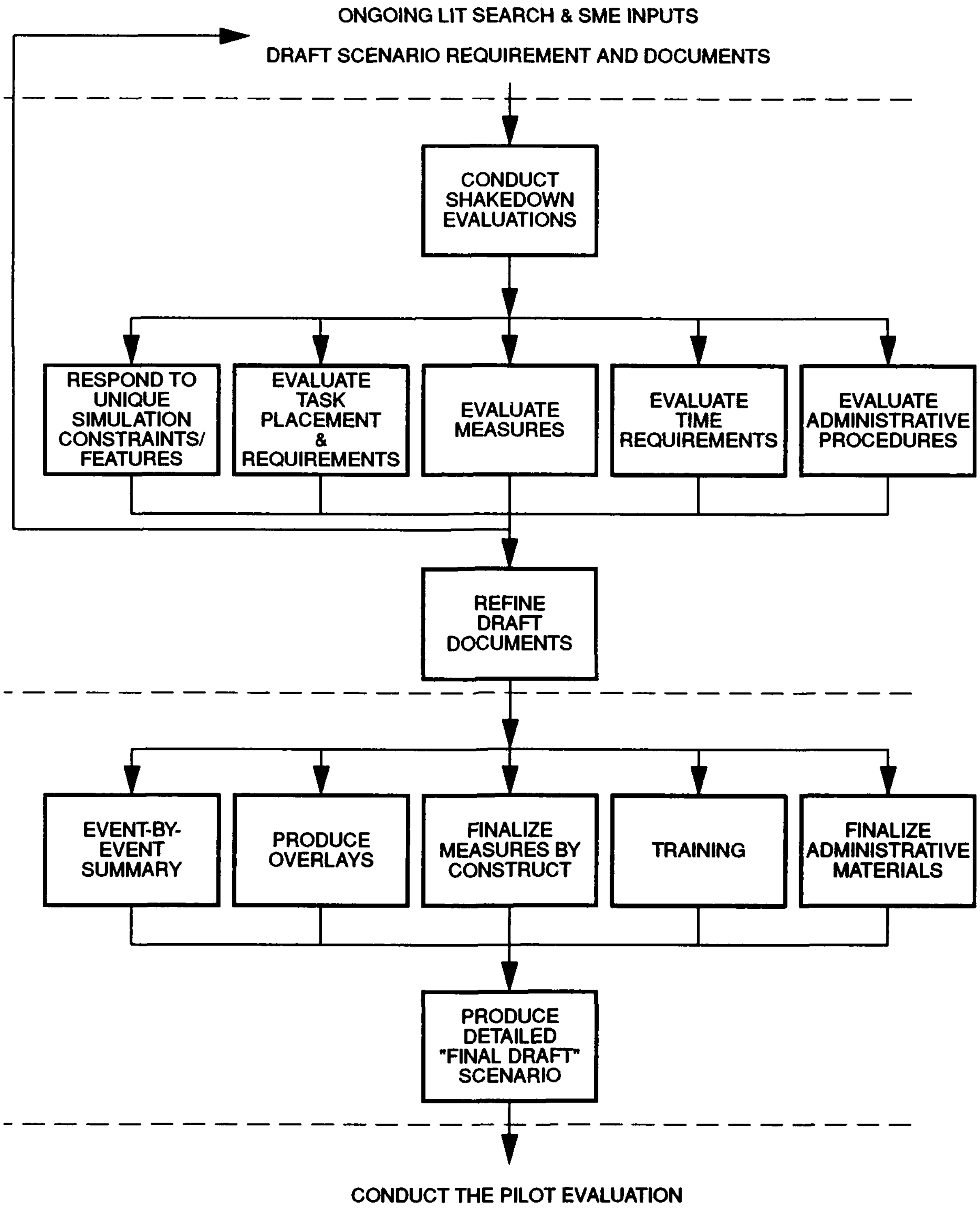

Figure 7. Scenario development: steps for iteratively redesigning the scenario

Moreover, specific SIMNET-D resources require repeated use to learn. The operation of the SAFOR to simulate friendly and enemy vehicles requires much practice and procedure refinement. SAFOR vehicle routes and vehicle parameters must be identified based upon an iterative refinement process.

Specific measurement capabilities and requirements can also affect scenario design and requirement issues. To capture some measures, evaluators may have to tailor standard operating procedures. For example, to capture accurate own location information for assessing checkpoint arrival performance, administrators may have units stop when they reach checkpoints and report their location. Current operating procedures, however, require units only to report a checkpoint within 30 seconds of crossing it.

Outline Time Requirements. Based upon iterative scenario refinement, evaluators should outline the time requirements for mission and task execution. Evaluators should focus upon ensuring that the scenarios could be completed within time constraints. This does not mean that all units must complete the scenario within the allotted time, but only that an effectively performing unit would have more than enough time to fully execute the exercise. One very important measure of combat effectiveness can be the percent of missions completed by the units with the time allotted. This measure proved to be the greatest discriminator between units equipped with new systems and baseline units in the Block II POSNAV and IVIS evaluations.

Evaluate Administration Requirements. While iteratively refining scenario design, evaluators should also begin to finalize administration procedures. Controller scripts, used by exercise administrators and trainers to coordinate and initiate scenario events and any non-automated performance measurement, should be drafted and refined. These scripts should indicate necessary administrator actions and communications, an event-by-event exercise description, data collection forms, and event initiation requirements (e.g., vehicle and bombing locations, and timing instructions). Evaluators should be concerned with ensuring exercise standardization and accurate performance assessment. Administrator requirements should be shared by several support personnel, if necessary, to ensure that controller workload is not too high. Backup measurement capabilities, if available, should also be implemented. For example, in the IVIS evaluation, data collection forms used for collecting platoon communications data were complemented by automated radio traffic recording to protect against missing data.

Evaluate Measures. At this stage, evaluators should be able to review sample printouts for automated performance measures, as well as the assumptions inherent in any programming routines. Evaluators should carefully review these automated measures and procedures to ensure that the right measures are being captured reliably. Specific performances can also be simulated to evaluate programming accuracy and consistency. Furthermore, when possible, hand calculated estimates should be compared with analyst program estimates. Specific measurement verification routines should also be identified to ensure data accuracy. These verification procedures span several requirements. Procedures should be developed to evaluate the accuracy of any data transfers (e.g., from log sheet to PC database, from printout to PC database, from PC database to measure generation, and from simulation to printout). Of particular concern are occasional aberrant values (outliers) which inadvertently appear in the simulation. These outliers can produce values which deviate greatly from the majority of values and, if undetected, can result in data unreliability and false conclusions. Data analysis and verification procedures should not be left to SIMNET-D analysts, but should be shared across support personnel. In fact, for the POSNAV and IVIS evaluations, all data transfers were independently evaluated by at least two researchers.

Feedback. Evaluators should ask all test support personnel, SMEs, and soldiers who participate in the draft scenario refinement process for feedback as often as possible. All participant comments and suggestions should be reviewed and a quick response procedure should be used to refine scenarios and associated training, administration, and assessment materials. Evaluators should expect to have to frequently stop and restart scenarios and other procedures to make changes and ask for feedback. For example, soldiers may miss targets or respond in an unexpected manner to battlefield events.

Produce the Detailed "Final" Draft Scenario

The repeated shakedown evaluation which are part of the draft scenario redesign process should result in significant feedback for scenario adjustments. Some of these adjustments are reflected in the draft scenario finalization process that was shown earlier in Figure 7.

Event by Event Summary. As a result of the iterative scenario redesign process, evaluators should document in detail the tactical events and requirements which comprise the "final" or pilot ready scenarios. In addition to task information, this log should include controller requirements and instructions, simulator, target, SAFOR, PVD, and MCC initialization and placement parameters, a complete operations order (if applicable) or soldier handout materials and resources.

The log should be as thorough as possible. The level of detail should enable individuals unfamiliar with the scenario to reproduce the scenario to precise detail. Moreover, the log should be fully briefed and forwarded to all test support personnel as part of the training.

Produce Overlays. "Final" scenario map overlays should be developed and copies should be produced for the pilot evaluation. Several sets of overlays may be necessary, including a complete set of overlays which depict all tactical situations, tasks, and vehicle, bombing, and control point locations on the SIMNET-D and PVD maps for scenario administrators. A less detailed set of standard overlays should be prepared for soldier participants. In many cases, in-tank data collectors should be completely knowledgeable of the scenario and events. By following along in-tank with a completely annotated set of overlays, data collectors can more objectively and accurately assess crew behavior and discover exercise problems during the pilot evaluation.

Outline Measures by Construct. To organize and evaluate final measurement selections and procedures, measures should be fully documented. This documentation should include the overall measurement constructs (e.g., gunnery), the component measures (e.g., opening time), and all specific measure definitions (e.g., the time in seconds from target identification via intercom communication to first round fired at target).

Detail is especially important during this stage. For example, further detail on the above opening time definition may be necessary to delineate what is counted as a first round fired at target or how the intercom target identification command is recorded. This documentation is especially important for SME and soldier review during subsequent pilot evaluations.

Training. "Final" training materials and administration procedures should be fully documented. Procedures and requirements should be fully briefed to all test support personnel. The importance of following scripts, documenting problems, and ensuring trainee learning should be stressed. Specific training procedures should identify requirements for training task repetitions and after action reviews. For example, trainers in the POSNAV and IVIS evaluations were required to document soldier success on a training task at least three times in a row before they could move to the next task. In addition, sample tasks were included for the trainers for the purpose of assessing soldier performance.

After action review requirements were also addressed, including the feedback responsibilities of in-tank observers and out-of-tank exercise controllers. For example, POSNAV trainers gave soldiers specific feedback on their use of in-tank performance resources (e.g., maps, POSNAV, radio, vision blocks and sights). Exercise controllers, using data collection logs and analysis printouts (if available), described more specific objective task performance information such as report accuracy and timeliness and obstacle bypass accuracy.

Conduct the Pilot Evaluation

The scenarios should never be considered complete until they have been evaluated in a full-scale pilot evaluation setting. No evaluation should be performed without a preceding pilot effort. A pilot evaluation provides evaluators with an opportunity to conduct a full-scale "dress rehearsal". As a result, evaluators increase the probability that they have identified all undetected scenario design, administration, training, and assessment problems before the actual evaluation. The pilot evaluation process is shown in Figure 8 and is described below.

Pilot Evaluation Design. The pilot evaluation should be conducted using at least two unique groups of soldiers. If one is conducting a new system evaluation, evaluators should try to evaluate as many treatment conditions as possible. For example, in the POSNAV, IVIS, and CITV evaluations, unique groups of soldiers participated in the pilot to examine both the baseline and experimental (new system equipped) conditions. The use of two groups of soldiers provides the evaluators an opportunity to conduct one pilot evaluation using one group, then to redesign and repilot the scenarios with the second group.

The pilot soldier participants should be representative of the soldiers expected in the actual evaluation. Moreover, troop time requirements should be extended beyond expected evaluation times to ensure that sufficient time exists for the collection of soldier feedback in individual or group discussion sessions.

Although the pilot evaluation should be considered a full dress rehearsal, evaluators should not adhere strictly to evaluation procedures. The participants should be informed, up front, of the shakedown nature of the evaluation. Evaluation objectives, scenario design efforts, and other requirements should be fully explained to the soldier participants. The soldiers should be considered subject matter experts who will provide additional suggestions for improving all areas of the scenario process. Soldiers should be encouraged to offer criticism when necessary.

To effectively respond to soldier and SME team comments and suggestions, evaluators should also maintain a quick response capability to adopt scenario-related changes rapidly. The soldier, if possible, should see and evaluate the results of their comments.

Pilot Requirements. Several areas of scenario design, execution, and assessment should be examined as part of the pilot evaluation. These include: (a) scenario design and administration; (b) data collection and performance assessment procedures; (c) training design, administration, and evaluation procedures; and (d) equipment reliability. These reviews can be supported by feedback from one's SME team, soldier participants, and test support personnel, as well as actual performance data.

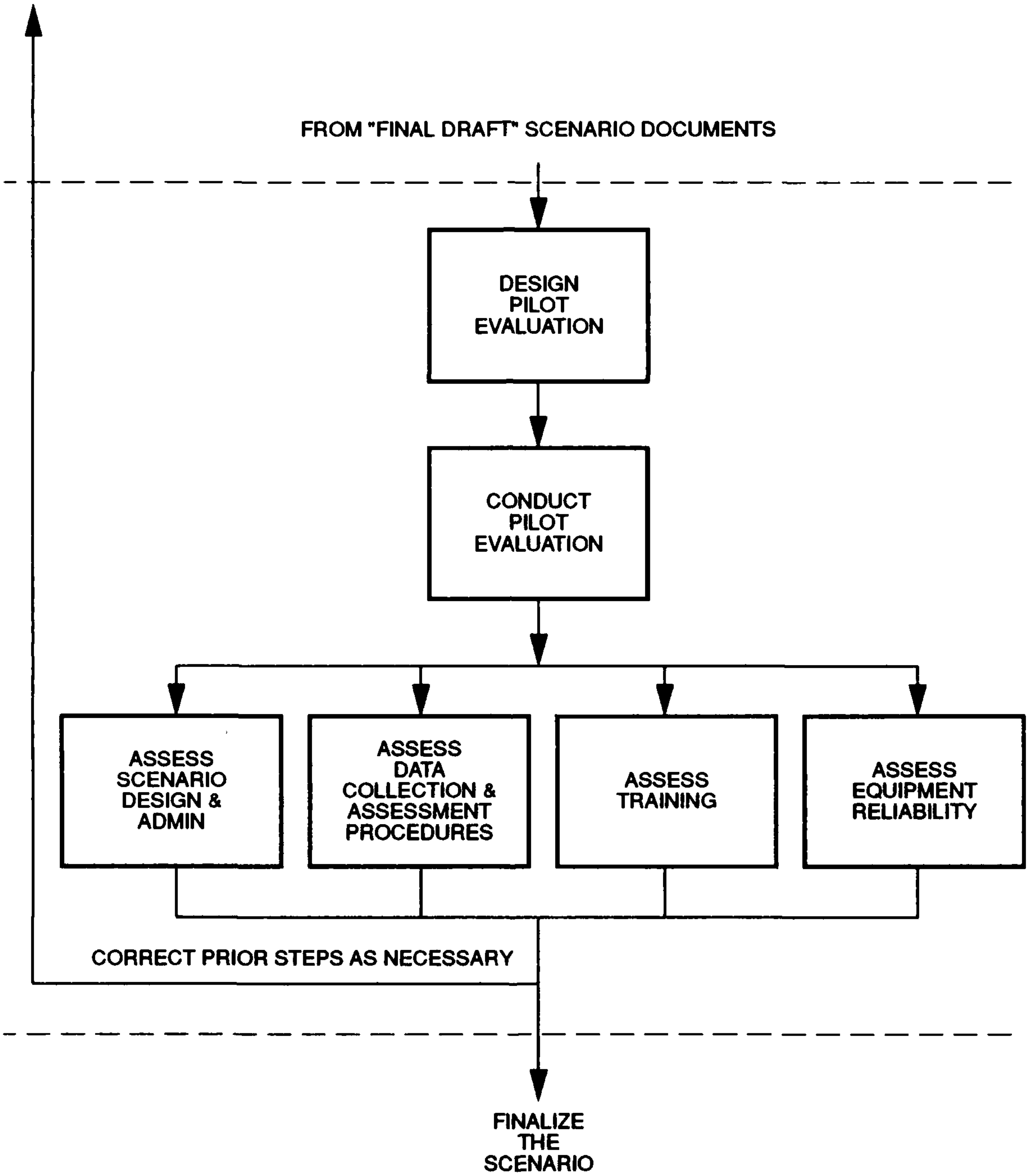

Figure 8. Scenario development: steps for conducting the pilot evaluation.

Scenario Design and Administration. Despite the numerous dry runs and simulation runs that should have been a part of the draft scenario design process, evaluators will likely discover problems with their scenario events during the pilot evaluation. These problems typically range from major problems with task placements and requirements to minor changes in administrator communications.

The pilot evaluation provides an opportunity to realistically assess the saliency and appropriateness of exercise tasks and the actions which prompt or require them. Evaluators may detect unanticipated soldier reactions to scenario requirements. Moreover, soldiers may use different strategies to complete the scenario. For example, some units may exploit available artillery support in a combat mission while others may not. The pilot evaluation provides an opportunity to assess the ability of the scenario execution and measurement process to respond to such performance variability, especially across experienced and inexperienced, poor and excellent performing, units.

Data Collection and Assessment Procedures. The pilot evaluation may provide the first opportunity to assess the data collection, measurement, verification, and analysis requirements. Evaluators should collect all data that would be gathered in the actual evaluation. Evaluators should examine the time it takes to collect and verify data and the workload of collectors and administrators. Moreover, the need for additional measures or different measurement definitions not anticipated could result from soldier, SME, test administrator, and trainer feedback, as well as from the actual performance data.

Training. The soldier training conducted as part of the pilot evaluation will give the evaluator an opportunity to examine the effectiveness of procedures used in advance of the actual exercise. Evaluators should examine and document all training feedback and performance data. Evaluators should also look at the time it takes to conduct the training, as well as the effectiveness of training program phases and support materials.

Equipment Reliability. One of the greatest benefits of the pilot evaluation, at least in the POSNAV, IVIS, and CITV evaluations, is that it provides a chance to continually use and evaluate the SIMNET-D equipment. Regardless of the efforts expended during scenario design, evaluators will probably experience frequent equipment breakdowns during the pilot. Evaluators should be prepared to document all breakdowns and forward them immediately to SIMNET-D engineers. Equipment particularly sensitive to breakdowns are the simulators, any new combat developments, and the SAFOR.

Finalize the Scenario

Evaluators should schedule at least a two week break between the pilot and actual evaluations to respond to the information gained during the pilot. This time is required to finalize all important scenario documents and procedures, including scenario design and administration procedures, data collection and assessment routines, training design, and administration and evaluation materials. Moreover, this phase of scenario development provides the evaluators with an opportunity to document and review the limitations of the current effort. Scenario finalization requirements are described below and are illustrated in Figure 9.

Scenario Design and Administration. Based on the information gleaned from the pilot evaluation, evaluators should make all necessary changes to the scenarios during this phase. Final scenario overlays, event logs, controller scripts, and administration procedures should be produced and forwarded to all test support personnel. Materials should also be copied and stored for the soldier participants. All materials should be reviewed in a group setting with all test support personnel and the SME team.

Data Collection and Administration. All assessment procedure changes that affect current automated measurement routines should be forwarded to SIMNET-D analysts immediately following the pilot. Final assessment schedules should be created, including specific requirements for data formats and transfer. Final data collection and verification assignments should be reviewed. All data collection forms, measure documentation, and assessment procedures should be reviewed in a group setting with all test support personnel and the SME team.

Training Materials and Procedures. Based upon pilot evaluation findings and feedback, evaluators should reevaluate the soldier and test support personnel training needs. Any additional training requirements should result in the generation of more training materials and procedures. A final training package, including administration requirements, evaluation materials, and other resources, should be prepared and reviewed with all test support personnel and the SME team.

Documenting Evaluation Limitations. Once the pilot evaluation is complete and all evaluation materials are revised and documented, evaluators should reevaluate and document the limitations of their effort. Evaluators should be objective in outlining their final concerns and be prepared to document these concerns in any reports or briefings which result from the evaluation. The overall rationale for the final evaluation design, including training and assessment practices, should be documented in preparation for the final report. Test support personnel, if possible, should participate in this documentation, especially where it is related to their specific contributions to

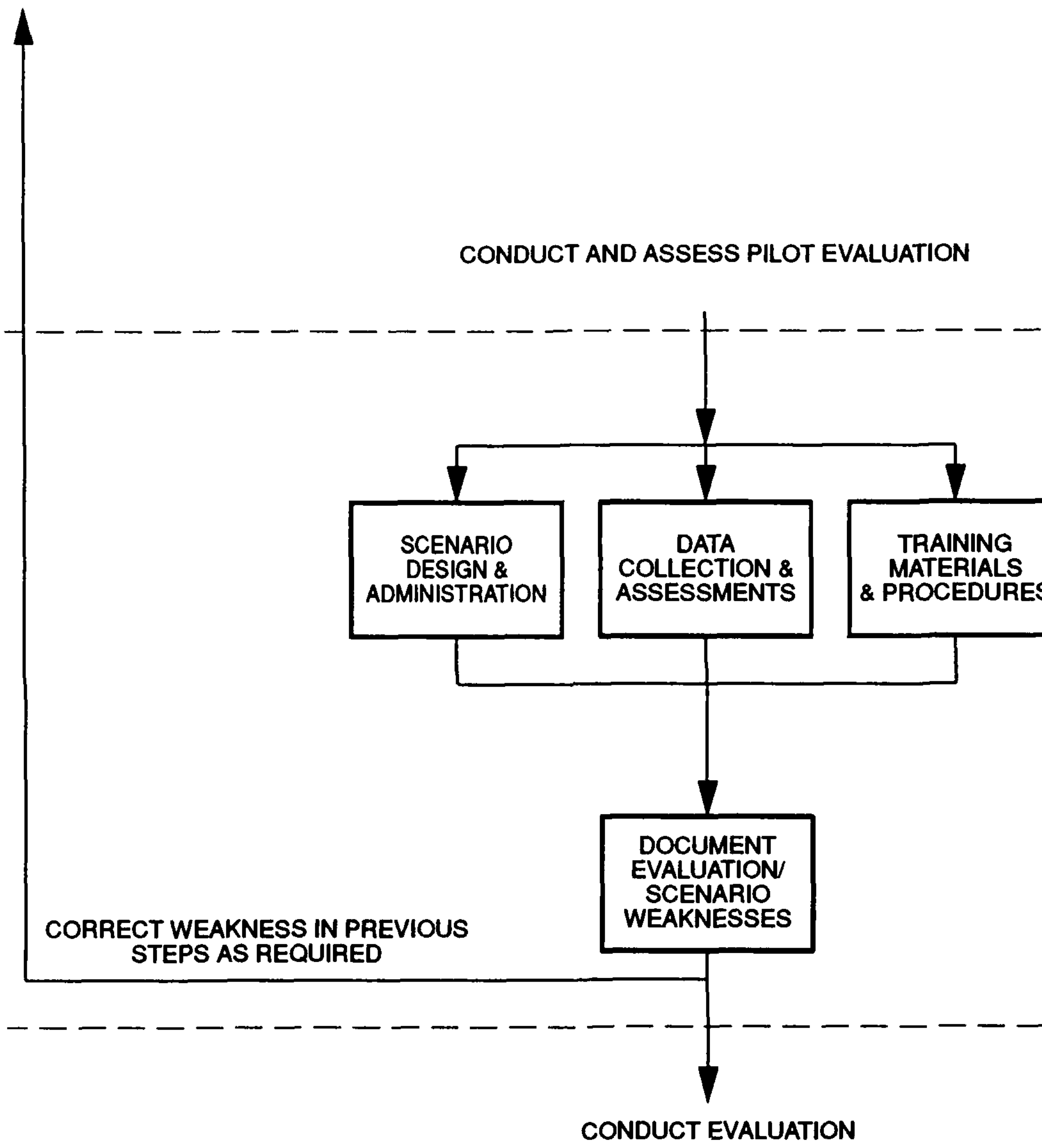

Figure 9. Scenario development: steps for finalizing the scenario.

the scenario development effort. If necessary, evaluators should be prepared to make any required revisions to scenario materials to reduce the effects of any limitations.

Summary and Conclusions

In this report, systematic steps and process considerations are outlined for the development of training and evaluation scenarios using SIMNET-D. The guidelines are set up so that an evaluation team can use them to work together and prepare to conduct an evaluation using SIMNET-D. This evaluation team, it is assumed, would consist of a combat development action officer, experimental researchers, subject matter experts, and scenario developers.

The seven step process described in this report emphasizes deriving, documenting, and verifying scenarios. Scenarios for concept evaluations must be verified in conjunction with the research objectives, the research design, measurement considerations, and data analysis and interpretation considerations. In addition, the scenarios must be compatible with the capabilities and limitations of SIMNET-D. The major steps in preparing scenarios are:

1. Define the objectives of the research and/or training.

2. Select the evaluation and scenario strategy.

3. Draft the general scenario requirements.

4. Iteratively redesign the scenario.

5. Produce the detailed "final" draft scenario.

6. Conduct the pilot evaluation.

7. Finalize the scenario.

By following these major steps, the evaluators will assure that several common, but disastrous errors of omission are avoided. These steps assure that:

* Sufficient emphasis is placed upon one's purposes and objectives in scenario planning.

* An SME team (not just a single SME) contributes to and complements the evaluation team throughout the preparation.

* The essential implications for the scenarios are systematically documented during their development.

* Scenarios and other interacting aspects of the evaluation are iteratively designed and refined using critical shakedown and pilot evaluations as vital verifications.

* Data collection and analysis routines are thoroughly verified and documented.

For each major step, detailed process considerations which guide the evaluation team toward a successful and purposeful evaluation are offered. These guideline considerations are not substitutes for the expertise which the action officers, evaluators, SMEs, or scenario developers bring to bear upon the investigation. Nor do the detailed guidelines replace more fundamental reference sources on SIMNET functions, on combat or training operations, on task requirements, or on experimental design. Rather, when followed, they systematically focus and integrate efforts toward tailored scenario treatments that address the research objectives. These detailed process considerations place emphasis on:

* Detailed but simple documentation that allows one to trace one's considerations from objectives through changes to final documents. This documentation also promotes consistent control and standardization for the conduct of the scenario-based evaluation.

* Systematically, yet at times, simultaneously integrating considerations related to major steps in scenario development.

* Using extensive task and performance data that may be available. Likewise, drawing upon information and lessons learned in previous research associated with the areas to be considered.

* Attending to the staff and participant training demands of the scenarios, the simulation, and the evaluation.

* Using SMEs throughout the scenario development.

* Interacting with the originators throughout.

* Creating measures tailored to specific scenarios and the research purposes rather than accepting the default measures which are easy to collect within the SIMNET system.

* Verifying the scenarios, the measures, and their analytic manipulations during the shakedown and pilot evaluations.

* Documenting the weaknesses of the evaluation including those associated with simulation and scenario shortcomings.

SIMNET-D is a sophisticated evaluation tool. These guidelines provide a straight forward means to deal with this complexity and to prepare for a successful investigation.

References

Barron, R. C., Lutz, W. G., Degelo, G. J., Havens, J. W., Talley, J. W., Smith, J. R., & Walters, R. F. (1976). Degradation of tank effectiveness (Test Report No. 325). Fort Hood, TX: Training and Doctrine Command Combined Arms Test Activity.

Bolt, Beranek, & Newman Laboratories Incorporated (1986). Developmental SIMNET data handbook. Cambridge, MA: Author.

Chung, J. W., Dickens, A. R., O'Toole, B. P., & Chiang, C. J. (1988). SIMNET M1 Abrams main battle tank simulation: Software description and documentation [Revision 1] (Report No. 6623). Cambridge, MA: Bolt, Beranek, & Newman Laboratories Incorporated.

Du Bois, R.S., & Smith, P.G. (in preparation). Simulation-based assessment of automated command, control, and communication capabilities for Armor crews and platoons: The Intervehicular Information System (ARI Technical Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Du Bois, R.S. (in preparation). Instruments for evaluating the Intervehicular Information System (ARI Research Note). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Du Bois, R.S. (1989). Simulation-based command, control, and communication exercise for Armor small unit commanders (ARI Technical Report No. 866). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Du Bois, R.S., & Smith, P.G. (1989). A simulation-based evaluation of a Position Navigation system for Armor: Soldier performance, training, and functional requirements (ARI Technical Report No. 834). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Gound, D., & Schwab, J. (1988). Concept evaluation program of Simulation Networking (SIMNET) (Technical Report No. 86-CEP-0345). Fort Knox, KY: U.S. Armor and Engineer Board.

Keppel, G. (1982). Design & Analysis: A Researcher's Handbook. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Miller, D.C., & Chung, J.W. (1987). SIMNET-D capabilities and overview. Cambridge, MA: Bolt, Beranek, and Newman Laboratories Incorporated.

Perceptronics Incorporated. (1986). SIMNET master documentation. Woodland Hills, CA: Training and Simulation Division.

Quinkert, K.A. (in preparation). Crew performance associated with the simulation of the Commander's Independent Thermal Viewer (CITV) (ARI Technical Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

U.S. Army Armor School (1987). M1 SIMNET Operator's Guide. Fort Knox, KY: U.S. Army Armor Center.

Winer, B.J. (1968). Statistical principles in experimental design. New York: McGraw-Hill Publishing Company.